



June 7, 1995

**Emissions Inventory Data Sharing**  
**EIIP Data Transfer Survey Results & Analyses**

Prepared for:  
Office of Air Quality Planning & Standards (OAQPS)  
Emissions Inventory Improvement Program/Data Management  
Committee

EPA Contract No. 68-W3-0020  
Delivery Order 5001-009

Technology Planning and Management Corporation  
Headquarters Park  
Beta Building/Suite 220  
2222 Highway 54  
Durham, North Carolina 27713



# TABLE OF CONTENTS

<b>INTRODUCTION</b>	<b>1</b>
<b>DOCUMENT METHODOLOGY</b>	<b>1</b>
<b>SCOPE</b>	<b>1</b>
<b>SURVEY BACKGROUND</b>	<b>2</b>
<b>SURVEY RESULTS</b>	<b>3</b>
Survey Form	3
Organizational Affiliation	3
Data Transfer Capabilities	4
Quantity and Frequency of Data Transfer	7
Data Sensitivity	10
Hardware	11
Future Resources	12
Preferences	13
Additional Comments	14
<b>OBSERVATIONS AND ANALYSES</b>	<b>15</b>
General Observations	15
<b>ANALYSES OF RESPONSES BY SURVEY SECTION</b>	<b>17</b>
Organizational Affiliation	17
Data Transfer Capabilities	17
Data Formats	18
Quantity and Frequency of Data Transfer	18
Data Sensitivity	20
Hardware	20
Future Resources	20
Preferences	21
<b>SUMMARY OF CONCLUSIONS: EIIP/DMC DATA TRANSFER SURVEY</b>	<b>22</b>
I. Media	22
II. Data Formats	22
III. Volumes and Frequency of Data Transmission	23
IV. Transmission and Communications	23
V. Hardware	23
VI. Application Characteristics	24

# **EIIP DATA TRANSFER SURVEY RESULTS AND ANALYSES**

**JUNE 1995**

## **INTRODUCTION**

The "EIIP Data Transfer Survey Results and Analyses" is a deliverable for Task 1 of Delivery Order 0009 of EPA Contract 68-W3-0020. It also serves as supporting information for the recommendations paper required by Task 2 of the delivery order. The Emissions Inventory Improvement Program (EIIP) Data Management Committee (DMC) approved this survey as a means of gauging the "user requirements" for data transfer within the emissions inventory program. The survey itself was designed to provide the DMC with data regarding the media and formats currently being used to transfer emissions data. The survey results, and their subsequent interpretation by the DMC, will lead to a recommendation regarding the base line medium to be used to transfer emissions data in the future. Initially, this will be done in the form of a "concept" paper to describe potential data transfer protocols. The concept paper will be used as a mechanism to solicit input from the EIIP community and will form the basis for a more formal and eventual recommendation.

## **DOCUMENT METHODOLOGY**

This document is being written using Microsoft Word 6.0, but is being distributed largely in WordPerfect 5.1 (DOS) format. The document includes a section with a version of the original survey form, with result counts or actual written comments edited into the form in their appropriate position. All returned surveys are available in printed form, regardless of the method used to return them (Internet, CHIEF, etc.). The original results were tabulated in either Microsoft Word 2.0 files or a Microsoft Excel 4.0 spreadsheet. The section of this document that displays survey results (i.e., the annotated "survey form") summarizes the results totals stored in an Excel 4.0 spreadsheet. The mailing lists used to distribute the announcement letter and the survey forms are available in a Microsoft Access 2.0 database file. Further spreadsheets were used to develop conclusions from the analysis of the survey results, and straw-man recommendations for the DMC are provided in the final section based upon the conclusions.

## **SCOPE**

The DMC's mission, as it is currently defined, is to provide the EIIP with three products that can be interconnected and used by emissions inventory participants to efficiently and accurately transfer emissions data between one another. These three products, when complete, should fit on top of each other ("interconnect") much like the layers of a cake. These "layers" are listed below in the order in top-down (logically) fashion:

- a data model
- a data transfer format
- a common data transfer medium

As is evident in this list, a common data transfer medium will provide the foundation upon which an emissions inventory data model and data transfer format can operate. A data model will enable emissions inventory participants to use the database product of their choice to create output files that share the same internal structure as all other database applications used by all other emissions inventory participants. Once this data model is available to participants, a data transfer format can be agreed upon that will allow emissions inventory data to encode, then transmit and receive, emissions data over the common data transfer medium. Appropriate tools and transfer applications will need to be chosen by the particular end-user to support the particular data transfer formats to be used. These tools may include both data transfer software and data translation software for format conversion. It is anticipated that these tools will be available commercially, but may need customization or special development by specific end-users to support the specific characteristics of the chosen EIIP/DMC format.

It is most important to emphasize that this document provides the basis for recommending a common data transfer medium for emissions inventory data. Constructing a data model and data transfer format are separate pursuits of the DMC. The analyses and recommendations contained herein are provided independently of the efforts underway regarding the data model. However, some EDI-related (electric data interchange) questions, specifically those regarding "record formats", were asked within the survey knowing that evidence of an especially popular record format *might* provide the DMC with a starting point for whatever data transfer format is eventually chosen.

## **SURVEY BACKGROUND**

The intent of this survey was to query a representative sample of EPA, state, and local emissions inventory users regarding their current and future capabilities for transferring emissions inventory data, electronically or otherwise. We asked a number of questions regarding the frequency and quantity of emissions data transfer. We also attempted to determine the file and record formats used by emissions inventory users during data transfer, and individual *preferences* regarding formats and transfer methods. Of secondary concern was determining the general nature of the hardware and software currently used by those individuals who transfer emissions inventory data.

TPMC initially sent an "announcement letter" via U.S. Mail to approximately 160 individuals within a very broad range of EPA, state, and local environmental agencies, and to a few individuals in the commercial (or industry) sector with an active interest in emissions inventory. The mailing list was provided to TPMC by the EPA delivery order project officer. The mailing list consisted of a predominance of emissions inventory program participants within the states. The announcement letter requested that those individuals who were interested in participating in the survey reply to TPMC using a self-addressed stamped envelope. Of those individuals on the initial mailing list, 61 indicated that they were willing to participate in the survey.

The surveys were mailed using three methods and according to the stated desire of each respondent to the announcement letter: U.S. Mail, TTN mail (CHIEF BBS), and electronic mail via the Internet. The mailings were divided almost equally between each of the three methods. Addressees were asked to mail the surveys back no later than Friday, May 5, 1995,

which was approximately 2 weeks after a majority of the surveys were sent. Follow-up e-mail messages were sent to all Internet users who did not respond by May 5, and to all CHIEF users. In the interest of saving time, phone calls were placed to U.S. Mail participants where possible for those who had not responded by the deadline.

Thirty-nine completed surveys - approximately 64 % of the sixty-one individuals who expressed interest in participating - were returned to TPMC. Surveys sent via U.S. Mail were returned with the most consistency, followed by those sent via Internet. CHIEF did not prove a very successful medium through which to obtain responses (only 2 surveys initially sent via CHIEF were returned).

## **SURVEY RESULTS**

The results of the survey are provided below. Many questions required that respondents provide either a check-mark, a 'C' for "CURRENT CAPABILITY", or a 'P' for "PLANNED CAPABILITY" as appropriate for the question. A count of the total number of respondents who marked each part of a question is included in the underlined space provided on the initial survey form. All respondents reported on "CURRENT CAPABILITIES" but many did not respond on what was "PLANNED". In these cases, it was inferred that the "PLANNED" approach, at a minimum, is to continue the "CURRENT" approach. These inferred responses were added to the "PLANNED" totals.

Other questions required a textual response. These have been compiled into summary form in the space following each question. Comments that are exact quotes are enclosed in quotation marks and shown in italics. However, in the vast majority of cases below, only the related survey number or references to the names of individuals have been edited out of the original results. Some quotations have been logically grouped to provide the reader with a ready understanding of the breadth of the responses.

### **Survey Form**

#### ***Organizational Affiliation***

1. Please place a check - mark beside your organization's affiliation.

#### *Invitation letters and/or surveys sent out: (Total 222)*

Federal gov't:	<u>30</u> (14%)	State gov't:	<u>114</u> (51%)
Local gov't:	<u>73</u> (33%)	Commercial:	<u>5</u> (2%)
Other	<u>0</u>		

#### *Respondents: (Total 39 responses received or 18% of invitations)*

Federal gov't:	<u>6</u> (15%)	State gov't:	<u>24</u> (61%)
Local gov't:	<u>8</u> (20%)	Commercial:	<u>0</u>
Other:	<u>1</u> (2%)		

**Data Transfer Capabilities****Media Type**

2. Please place a check-mark(s) in the 'C' column next to the medium (or media) that you currently use to transfer emissions inventory data file to emissions inventory participants outside of your organization. Please place a check-mark(s) in the 'P' column beside the medium (or media) that you plan to use during the next 2 years. Check as many as necessary:

<u>Media Type</u>	<u>C</u>	<u>P</u>
Paper	34	24
Floppy Disks	30	26
Magnetic Tape	10	11
Modem	26	29
Direct Network Access*	17	21
Other _____	2	2
N/A	1	1

3. If you are using (or planning to use) magnetic tape, please list size and format:

**Size:** (1/4" cartridge, 8mm, 4mm, 9 track, etc.)

Current: \_\_\_\_ (includes: "MiniData Cartridge 120 Meg, 307.5 ft", "1/4 in. cartridge [2], 8mm[2], 9 track [4], 4mm)

Planned:\_\_\_\_ (includes: 1/4" cartridge, 8mm [2], 9 track, Maynard Mainstream MS2200)

**Format** (QIC 80, tar, VMS saveset, etc.)

Current: \_\_\_\_ (includes: 3M, DC2120, QIC 80[3], tar, IBM , ASCII/EBCDIC, 20G digital cartridge)

Planned:\_\_\_\_ (includes: QIC 80, tar, ASCII/EBCDIC)

4. If you have a modem, please place a 'C' beside its current baud rating and a 'P' beside the baud rating to which you plan to upgrade within 2 years:

**Current:**

1200	<u>0</u>	2400	<u>1</u>	9600	<u>17</u>	14.4Kbps	<u>22</u>
28.8Kbps	<u>2</u>	Other	<u>2</u>	N/A	<u>2</u>		

**Future:**

1200	<u>0</u>	2400	<u>1</u>	9600	<u>17</u>	14.4Kbps	<u>12</u>
28.8Kbps	<u>20</u>	Other	<u>2</u>	N/A	<u>2</u>		

5. If you use file transfer software with your modem, please put a 'C' beside the product that you use and a 'P' beside the product to which you plan to upgrade within 2 years:

Kermit	C: <u>9</u>	P: <u>9</u>
ProComm	C: <u>23</u>	P: <u>21</u>
Qmodem	C: <u>2</u>	P: <u>1</u>
Crosstalk	C: <u>12</u>	P: <u>11</u>
MS-Windows	C: <u>0</u>	P: <u>4</u>
Other <sup>†</sup>	C: <u>11</u>	P: <u>9</u>
N/A	C: <u>0</u>	P: <u>1</u>

<sup>†</sup>Other - Current: Includes: "IBM", Chameleon, Banyan, "C IRMA", Quick Link 2, Rapid Filer, SNA/SAA, MS Win RAS (assumed to be Windows Remote Access), Microlink, Zmodem

<sup>†</sup>Other - Future: Includes: Chameleon, "possible BBS", Rapid Filer.

6. Many organizations now have, or are planning to purchase, access to the Internet or a Value-Added Network (VAN). If you currently have direct access to the Internet or one of the VANs listed in the table below (i.e., you can connect to one of these services directly from your PC), please place a check-mark(s) of the column marked 'C' within the appropriate row. If you plan to purchase access to one or more of them, please place a check-mark(s) in the 'P' column in the appropriate row.

#### **The Internet and Value-Added Networks**

<u>Internet/VAN</u>	<u>C</u>	<u>P</u>
Internet	19	29
CompuServe	2	3
America Online	0	0
Prodigy	1	1
MCI/Tymnet	0	0
Sprint	0	0
EPA's NDPD SNA*	14	6
Other <sup>†</sup> _____	4	3
N/A	2	2

<sup>†</sup>Other - Current: "AIRS-EPA", "Merit", "Chief", "InfoAmerica"

<sup>†</sup>Other - Future: "T1 to State", "Merit", "InfoAmerica"

7. Please place a 'C' next to "FTP" if you currently have an Internet connection and have software that will transfer files using FTP. Place a 'P' next to "FTP" if you do not yet have FTP, but plan to have it within the next two years. Using the same rules, please place a 'C' or a 'P' next to "Mosaic" if you have, or are planning to have, direct access to the World Wide Web. (Please see the "On Line Help File" for an explanation of these terms):

FTP C: 19 P: 24  
Mosaic C: 13 P: 25

### Data Formats

8. When you exchange your emissions data with outside organizations, which of the following formats (if any) do you use (please place a check-mark beside all that apply). See the "On Line Help File" for an explanation of data formats):

<i>FILE</i>	ASCII	<u>34</u> (DOS, UNIX, VMS etc., files)
<i>TYPE</i>	EBCDIC	<u>8</u> (IBM mainframe files)
	Binary	<u>7</u> (special and unique to each computer type)
	Other	<u>2</u> ("text and dbf", "DAT tapes")
	N/A	<u>1</u> ("not decided yet")
<i>FILE</i>	Spreadsheet	<u>21</u>
<i>FORMAT</i>	Type: (Lotus, QuatroPro, Excel, FoxPro, dBase)	
	WordProcessor	<u>22</u>
	Type: (WordPerfect and Word)	
	Generic Text File	<u>14</u>
	Other:	<u>8</u> (includes "flat file", "Paradox"[3], MS Access[2], Foxpro[5], dBase[6])
	N/A	<u>0</u>
<i>RECORD</i>	ESF	<u>0</u>
<i>FORMAT</i>	AIRS	<u>25</u>
	EPS-input	<u>10</u>
	EPS-output	<u>6</u>
	X12	<u>0</u>
	Other <sup>†</sup>	<u>12</u>
	N/A	<u>1</u>

<sup>†</sup>Other - (incl. "flat file", "gemap, rapids", "columnar format", "NAPAP, INT EL", "fixed format", "dbf, db", "Quicreports", "fixed length", "EGADS/AFS structure", "depends on request", "UNIX CU")



***Quantity and Frequency of Data Transfer***

9. When you transfer emissions inventory data, on average, how much data do you transfer (in megabytes)?

**Includes:**

*1 to 2 Meg*  
*0.5 meg*  
*<= 2 Meg*  
*0.5 - 1.5 Meg*  
*"N/A. We currently use AIRS directly. Information is there for them to download."*  
*3 Mb*  
*4-5 Meg*  
*2.5 Meg*  
*"unknown" [2]*  
*"1 Meg" [3]*  
*1.1 Meg*  
*"VARIES - OFTEN - 50-100 Mb*  
*AT A TIME - SOMETIMES 10 - 20 Mb*  
*Very Frequently - 1 - 8 Mb"*  
*"two- three meg."*  
*4 - 7 Meg*  
*1-2 Mb*  
*"no average"*  
*" < 2"*  
*" 10s to hundreds"*  
*10Mb*  
*"don't do much"*  
*"1M and up to 200M"*  
*".3 to .5MB"*  
*"Varies greatly - typically small < 1MB.*  
*Largest is transferring inventory to AIRS, approx. 70MB.*  
*"Difficult to estimate an average, depends on application."*  
*"40K"*  
*"100Mb"*  
*"100K compressed, uncompressed 6M"*  
*"1 Mb"*  
*"~3 is the largest"*

10. If there is no 'average' (i.e., you are required to respond to many *ad hoc* requests for emissions data, with resultant data files that vary wildly in size), please provide a short description of the range of data file sizes that you are required to share with other organizations:

**Includes:**

*“.5 Kb - 5Mb”*

*" none have been requested to this point"*

*"unknown - we currently transfer paper copies".*

*"Data file has over 3000 records , data files transferred range from 590 to 6000 + records."*

*" Projects range from:*

*(1) Development of regional multi-county photochemical modeling -GEMAP data files for Lake Michigan Ozone study and SE Michigan Ozone study.*

*(2) Ad hoc data queries for companies or consultants."*

*"1 Meg to 200 Meg”*

*“.25 Mb to 2 or 3 Mb”*

*"Anywhere from a short table in word processing format to dBase files of several hundred Megabytes when transferring inventory data for the entire United States."*

*“1 to 10Mb”*

*“Few hundred thousand bytes for a database”*

*"Some times we transfer small files and other times we transfer files of 50 Meg and will transfer up to 200 Meg."*

*“Data requests are usually name, address, plant emissions that meet some specified criteria.”*

*[response appears to be a question mark]*

*“Emissions data from model outputs or emissions measurement activities may range from 100K to over a gigabyte.”*

*“near 0 to 200Mb”*

*“from 3Kb to 8Mb”*

11. How often, *on average*, do you transfer emissions inventory data?

\_\_\_\_\_ number of times per \_\_\_\_\_

**Includes:**

*25 times a year*

*10 times per year*

*4-5 times per year*

*2 times per year*

*1 to 2 times per year*

*1 time per year [3]*

*1 time per month (3)*

*1 or 2 times per month for large data transfer*

*2 times per month*

*3 times per month*

*3-4 times per month*

*6-8 times per month*

*8 times per month (2)*

*1 time per week [3]*

*2 times per week [2]*

*2 -10 times per week*

*3 times per week*

*10 times per week*

*1 time per day*

*"The only emission data transfer that's done is to the EPA-AFS once a year as a series of batch files."*

*"see question 12."*

12. If there is no 'average', please provide a short description of the frequency with which you are required to share emissions data:

**Includes:**

*On line upload to AIRS - 3 per year*

*Batch upload to AIRS - 1 per year*

*Floppy Disk Met Data - 4 per year"*

*"Data sent to AIRS sometimes on a daily basis. Other than AIRS, data transfers are mostly annual emission inventories. This will change with the TITLE V implementation."*

*"Maybe 2 or 3 times a year. We are a small county without a lot of big sources so the amount of data from just stationary sources (not mobile) is not great. Most requests are for stationary sources. (We mainly transfer Air Quality Data to AIRS and other interested parties)."*

*"As more states are involved in regional multi-state oxidant modeling studies (such as LMOS), there is the need to provide data via modem or magnetic tape. As more companies become involved in emissions trading, there is greater need for companies to access emissions data for other companies via a network system."*

*"Most of the data we share is to fill requests from consultants, local air districts or EPA. The amount and frequency we share varies widely, but we do get a lot of requests. I'm guessing at the 2 times/week average."*

*"Our agency supplies emissions (and other data) to applicants requiring the data to support air quality dispersion modeling analyses for their air quality permit applications and to the concerned public, if necessary. "*

*"Depends upon the data requested and required."*

*"Some times we transfer several large files on one day and other times only one or two files per day. We do a lot of work in this area."*

*"To AIRS - 1 time per year. For other requests, it is about 2 - 3 times per week."*

*"This is awkward to answer, depending on how you define "transfer" and the purpose for the information transfer. Informal exchange of data between cooperating researchers on a project occurs frequently (i.e., weekly, monthly, or even daily). Data will typically be exchanged on a relatively infrequent basis (quarterly-annually) for external or peer review purposes. Public release of data at the completion of a project occurs infrequently (annually or longer)."*

**Data Sensitivity**

13. Do you electronically transfer 'sensitive' data (see the On-Line Help file for a definition of sensitive data) to other emissions inventory participants (Y/N)? \_\_\_\_\_

Yes: 14 No: 24

## Hardware

14. Please indicate below the type of computer you use to transfer emissions inventory data. If you use a personal computer, please provide the commonly known name for the CPU chip and its processing speed (e.g., 286/12, 386/25, 486/66, etc.). If you use a Macintosh, please provide the commonly known model name (Mac Quadra, etc.). If you use a workstation, please provide the manufacturer's name and a commonly known model name or number (Sun SPARCstation, DEC VAXstation, DG Aviiion, IBM RS-6000, etc.). If you use a minicomputer or mainframe, please indicate it's manufacturer and commonly known model number.

### Includes:

386/25

*COMPAQ Deskpro 386/25*

"486"

486/33 [5]

486/33, 486/50, 486/66, *Pentium on order, HP755 Workstation to be used w/T1 line.*

486/33dx 16M RAM

486/33 Unisys A16

*"COMPAQ 486 based PC operating at 33 MHz acting as a workstation under a COMPAQ prosignia server. The server is connected to a state mainframe computer that allows access to EPA's AIRS system."*

*"IBM PC 486/33 (linked to IBM Mainframe & UNIX workstations through TCP/IP and LAN"*

"486/50"

486/66 [9]

*COMPAQ DeskPro 486/66*

486/66 or Pentium 90

486/66 through Internet

486/66 or 486/50 or Pentium - all DOS or Windows

486/66 at present - p100 w/in next 6 months

486/66 20MB RAM, shortly to upgrade to 75Mhz. Pentium

*"We use PC's (486DX2/66 & Pentium/60), Mac Quadra, & Sun SPARC 20"*

*"Compaq 560 deskpro XL"*

*"Intel (or compatible) P5-90"*

*"The Bureau of Air Quality uses PC's w/ 486/66 and Pentium 90's. Our emissions inventory data resides on AIRS. The EI section goes through the Bureau LAN to an agency WAN to the Agency Mainframe to the SNA gateway, to the NCC's IBM where AIRS resides."*

*IBM Mainframe - Novell workstation*

*IBM 3090 -200 VM -CMS 5.0 (mainframe)*

*DEC 5000, 386/486 PC's, COMPAQ Prosignia server w/Unix, IBM Mainframes (MVS)*

*Hewlett-Packard Model 9000 Series 735/125*

***Future Resources***

15. One of the goals of this survey is to determine if there is a baseline medium that can be used by *all EIIP participants* to transfer emissions data. The previous questions in this survey have tried to determine what data transfer capabilities your organization currently has, and what it has formally planned to procure. We would also like to know if your organization has the resources to further upgrade its data transfer capability *as a direct result of the EIIP*. Please place a check-mark beside the item(s) that your organization may be willing to procure to enhance your ability to transfer emissions data:

Dedicated phone lines for modems	<u>22</u>
High speed modems	<u>28</u>
New software for data transfer	<u>25</u>
New software for data formatting (e.g. X12 translators)	<u>14</u>
Faster PCs or workstations	<u>31</u>
Access to a VAN (esp. CompuServe, Prodigy, etc.)	<u>10</u>
Direct access to the Internet/ World Wide Web	<u>27</u>

## **Preferences**

16. Given the current staffing levels, budget, and quantity and frequency of emissions data transmission, what do you (personally) think is the most effective and efficient means of transferring emissions inventory data? In other words, what do you think is the best way for your organization to transfer your emissions data now and in the near future?

*"Direct network access."*

*"For now - Disk and Modem to AIRS - Future - Modem to AIRS, Internet"*

*"Central Clearing House - national or regional"*

*"Direct Access to the Internet/ WWW"*

*"AIRS"*

*"Paper, disks, bulletin board, Internet (modem)"*

*"Through our current gateway with Seattle EPA Region 10."*

*" Large data files should be transmitted through FTP. Extremely large files should use something like 1/4" cartridge or 8mm tape. Internet access should be means by which data are set up for FTP."*

*" File transfer software/modem directly to end user system or tape transmittal."*

*" 1, Internet 2, Floppies "*

*" Modems initially. Direct high speed Internet connection later ( with modem backup)."*

*"Diskette"*

*" Tape or Internet for large data requests, Internet or floppy for small data requests."*

*"Internet"*

*"Direct transfer with PKZip"*

*"For those in the RTP area we could use the LAN networks. If we ever get access to Internet like [an EPA employee] we may use it. [The EPA employee] says we want do what he can do. I have not seen it work personally."*

*"Via high speed modem that is in a standard format to minimize the requirement for end user data manipulation."*

*"Via disk now; near future Internet - direct network access."*

*"It depends on who is receiving the data and what telecommunications facilities they have. If they are a state user, then transfer to the IBM mainframe would be the best. If they are on Internet, then that would be the best method."*

*"Modem + Internet!"*

*"now - floppy disk; future preferred via network"*

*"Via the Internet, using FTP and linked WWW home page"*

*"Through modem with communications software with users without Internet access. Possibly through with those that have access, but this is probably the least likely methods of the two."*

*"Password encrypted files via diskette or modem for point source data transfer between state and industry"*

*Direct Network Access"*

17. Which of these formats, if any, do you feel is best suited to transferring your data?

<i>FILE</i>	ASCII	<u>33</u>	(DOS, UNIX, VMS etc., files)
<i>TYPE</i>	EBCDIC	<u>2</u>	(IBM mainframe files)
	Binary	<u>7</u>	(special and unique to each computer type)
	Other	<u>1</u>	("text and dbf")
	N/A	<u>0</u>	
<i>FILE</i>	Spreadsheet	<u>17</u>	
<i>FORMAT Type: (Excel, Lotus, QuattroPro, Foxpro, "dbf or dbase files")</i>			
	Word Processor	<u>16</u>	
<i>Type: (WP and Word roughly tied)</i>			
	Generic Text File	<u>13</u>	
	Other <sup>†</sup>	<u>9</u>	
	N/A	<u>1</u>	

<sup>†</sup>*Other (incl. "FoxPro", "state specific", "RTF", "database file", "dBase", "Paradox")*

<i>RECORD</i>	ESF	<u>0</u>
<i>FORMAT</i>	AIRS	<u>19</u>
	EPS-input	<u>8</u>
	EPS-output	<u>6</u>
	X12	<u>1</u>
	Other <sup>†</sup>	<u>15</u>
	N/A	<u>2</u>

<sup>†</sup>*Other - (incl FoxPro, "as needed", "state specific", "unknown", "columnar format", "NAPAP, INT EL", "fixed format", dBase, Paradox, "fixed length", "EGADS or quick reports", "none of these is preferred")*

### Additional Comments:

*Creates own record format most of the time*

*"re Questions 8-13: SC is currently a direct AIRS user. Other organizations can have direct access to our data. Consultants often request data from AIRS. We run quick look (ad hoc) reports and generate hard copy or make files that we down load using Kermit and ProComm to copy to floppy disks."*

*"There is a good chance the BAQ will purchase a commercial emission inventory product such as ISTEPS. Even then the data will be batch uploaded plant-by-plant to AIRS. Other agencies including EPA, will be referred there for SC EI data. Utilizing AIRS meets data sharing needs at no additional expense to state and locals.*

*Includes mention of "state specific flat files (TBD) " for the record format for his data.*

*" At present, I send St. Louis County Data to the state of Missouri, which is supposed to transfer it to EPA."*

*"EPA should consider using grant funds to assist states in acquiring the proper connectivity & hardware/software to get the capability to use FTP and high speed data transfer software."*



*" At the time this survey was received, our administration informed us of intent to terminate departmental use of their mainframe computer by the end of the year. The net result would be either to maintain emission inventory data on a client/server SQL system or secure alternative contract services for maintaining the existing mainframe services. This decision has not yet been made by our Division staff as to what option will be selected, thus impacting future data transfer capabilities. "*

*"I have very definite ideas about our needs. We are a very large user of emissions and air quality data and want to play a part in transfer methods. Thank you for the opportunity to comment."*

*"I am an EPA employee, so most of my data transfers are to the public. The states input all of their own data to the AIRS system. Most of my data requests are in-house or through FOIA requests."*

*"I hope EPA will not abandon existing file transfer protocols and formats, such as AIRS and EPS input and output files. It has taken us some time to understand and become proficient in these formats. Also, we are currently using the I-STEPS software package from PES, written in Foxpro (a dBase derivative) for our point source inventory that creates ASCII AIRS format records. We have spent a great deal of time and resources on the system."*

*"Presently, trying to allocate funds to upgrade PC in the office, dedicate line for network and software."*

*"ASCII files are OK for transmitting text & numeric data, binary files are needed for graphics and images. I have attempted to provide answers for this questionnaire that represent our specific workgroup, not just me personally. Between us, Mac and UNIX platforms are used."*

*"Allegheny County believes that each agency should maintain its current inventory in AIRS so that users can access it there & retrieve in any desired format. This relieves the agency of the task of continually preparing specialized inventories for individual users and transporting them in some desired form or format. At the same time EPA should be strongly urged to improve and maintain AIRS as a user-friendly national database available to all users."*

I would like to receive a copy of the results of the survey: 26

I would like to receive a copy of the subsequent data transfer recommendations: 25

## OBSERVATIONS AND ANALYSES

### General Observations

The initial response to the announcement letter seemed enthusiastic. Many respondents returned the announcement letter within a day or two of receiving it, so that, within the first week of mailing the letters, a majority of the total responses had been received at TPMC. Almost 50% of the addressees on the announcement letter mailing list *eventually* responded positively to our request to participate. Considering that the announcement letter appeared "out of the blue", and with no follow-up calls or e-mail messages, this initial response was encouraging. Subsequently almost two-thirds (64%) of those expressing a willingness to participate actually returned the survey.

Follow-up phone calls were required after the survey return deadline had passed in order to prompt individuals to complete the survey. Most of the individuals contacted indicated that

they had no "problems" with the survey, they just had not had the time to complete it. This leads us to believe that the survey was not, as originally feared, overladen with telecommunications jargon or computerese, but that the survey certainly took some time to complete. The announcement letter asked each respondent how he or she would prefer to receive the survey - via U.S. Mail, CHIEF, or the Internet - and was then prompted to provide an address appropriate to the preferred mailing method. Perhaps individuals did not read or fully understand those instructions, as evidenced by the number of individuals who filled in more than one address blank (in these cases, TPMC chose to send the survey by whatever *electronic* means was specified in the letter).

The almost total lack of response (only 2 out of over 20) from CHIEF users adds to this concern. One person provided us with an incorrect address on CHIEF, and another advised us, when we placed a follow-up call to him, that he hardly ever logged into CHIEF, so had not yet seen the survey. Federal employees were frequently represented on the CHIEF address list. We received the highest rate (and quickest response) of returned surveys from those individuals to whom we sent *both* the announcement letter and the survey using the same mailing medium - the U.S. Mail. U.S. Mail replies were received at an even higher rate than occurred with Internet addressees.

When we initially entered the survey results into our spreadsheet and word processing files, we noticed that several individuals responded with a check mark where a 'C' (for "currently using") or a 'P' (for "plan to use") was intended. We interpreted check marks in these cases as a 'C'. It was disturbing to note, however, that *many* individuals did not respond completely within the tables where response spaces for current and planned usage were arrayed in columns and rows (Media Type, Modem, and Internet/VAN). This suggests one of three possibilities:

- the respondent was not aware of future plans or had none, or
- the planned approach was to continue the current approach , or
- the instructions weren't clear or specific enough regarding the intent.

In any case, our ability to forecast future trends was hampered by the incomplete responses. Where respondents provided current but *no* planned responses (approximately 50% of respondents to most questions), it was inferred that the planned approach was to continue the current approach, and these inferred responses were added to the future totals. A review of these overall results does permit us to overcome this handicap. For instance, it is abundantly clear by the aggregate response to question 4 (regarding current and planned availability of modems) that a significant percentage of respondents plan to upgrade existing 14.4Kbps (and slower) modems to 28.8Kbps modems. We interpreted a response as being "significant" to our analyses if it received at least a quarter (10 or more) of the overall response count. Clearly, if half (19+) or more of the respondents said that they were currently using, or planned to use, some file transfer mechanism or another, then the DMC would be obliged to acknowledge that mechanism in any recommendation to the EIIP Steering Committee.

## ANALYSES OF RESPONSES BY SURVEY SECTION

### Organizational Affiliation

The level of responses from federal employees in general matched the level of invitations that were sent to federal employees (14% of invitations/surveys went to Federal employees and 15% of the responses were from Federal employees). We received a very satisfactory response from state personnel (51% of invitations went to State employees and they returned 61% of the responses). The response level from individuals within local government was less (33% of invitations went to Local Government employees and they returned 20% of responses). The response from "commercial" interests was roughly proportional to the low (2%) overall percentage that had originally indicated interest in participating in the survey (the single respondent in "Other" was from a commercial firm).

Please note that states were favored over federal, local, and industry in the initial mailings in an attempt to keep the scope of the survey within reasonable bounds. Therefore, the high rate of return from state participants was encouraging and valuable. The geographic distribution of the returned surveys represented a roughly equal split between states west and east of the Mississippi River. The states responding included: AK, AZ, CA, DE, FL, GA, IL, IN, MA, MD, MI, MN, MO, NC, NV, OR, PA, SC, TN, TX, UT, and WI.

### Data Transfer Capabilities

Almost all respondents included the use of paper or floppy diskettes to transfer emissions data. Only three of the respondents (who had high-speed modem transfer) indicated that they did not use paper or floppies. 87% of respondents currently use paper and 77% currently use floppies. In addition, 67% of respondents indicated the use of modems for data transfer, and 44% indicated Direct Network Access among their current capabilities. Although the table in question 2 (Media Type) was one in which almost half of the respondents provided no indication of future media usage, even after inferring future use from current users providing no future plans, the use of paper for data transfer still declined by 30%. The most accurate statement that can be made regarding the future of floppy disk transfers, modem transfers, and direct network access is that their planned usage in the future is "significant." Magnetic tape usage only verges on being significant (26%).

Individuals who currently use modems possess relatively modern and fast ones (split almost evenly between 9600 baud and 14.4Kbps) with a significant intention of upgrading to 28.8Kbps within two years. The most popular file transfer software used with modems is ProComm, followed by Crosstalk, although Crosstalk does not quite qualify as "significant" using the guidelines provided above. Overall, it is clear that the majority of modem users have access to at least one, if not all, of the *de facto* file transfer protocols on the market today - Kermit, Xmodem, Ymodem, or Zmodem.

Of the individuals using "direct network access", it is clear that the Internet and the EPA NDPD SNA network are the only significant wide area networks in use. The Value Added Networks (CompuServe, Prodigy, etc.) were virtually non-existent in the results. Although

Internet users clearly plan on keeping their connections, we are cautious in how we interpret the apparent drop in EPA NDPD SNA connections, especially since this table was one in which several respondents completed only the responses in the "current" column. While the planned NDPD connections drop by eight (a 57% drop), the planned Internet connections increase of 10 (52%) only holds true based upon inferred plans for those who had Internet as a current connection but had not provided a future response. Taking a conservative view of the data, we believe the downward trend in NDPD SNA connections reflects the state of transition (and to some extent, uncertainty) in which both the NDPD out-year budgets and the AIRS database currently exist.

Of the Internet users, all indicated that they had FTP capabilities, and two-thirds indicated that they had access to the World-Wide Web ("Mosaic"). Almost all planned Internet users (except four) indicate that they will have access to the Web within the next two years.

### **Data Formats**

A very high percentage (87%) of respondents use ASCII files in their emissions inventory transfers. No other file type could be considered significant. This preference for ASCII files was reflected in the use of "generic text file" formats by over a third (34%) of the respondents. However, spreadsheets (split almost equally between Lotus and Excel with some Quatro Pro) and word processors (split almost equally between Word and WordPerfect) are used by over half the respondents. Overall, the responses to this group of questions suggests that a wide range of file formats are being used by individual groups. We noted that many individuals indicated that they had *both* Lotus and Excel or Quatro Pro (Lotus-compatible), or *both* Word and WordPerfect.

One category of file format that was represented within the responses - *PC-based database files* - had been inadvertently left out of the original survey form. As a result, we believe that the number of "write-ins" for database file formats should be considered significant. Popular data base formats that were indicated are primarily dBase derivatives or user-oriented relational databases including dBase, Paradox, MS Access, and Foxpro.

Record format responses revolved, for the most part, around AIRS (64%), "Other" (31%), and EPS-input (26%). We believe that this mix reflects the current range of applications that are used to store or process emissions data, with AIRS being the most predominant. However, the wide range of formats in the "Other" category indicates that there are a number of other applications (mostly PC databases) that are storing or manipulating emissions data. Although the total number of "Other" responses barely qualifies as "significant", we must assign a definite importance to them since they are all write-ins.

### **Quantity and Frequency of Data Transfer**

Questions regarding the quantity and frequency of data transfer must be asked whenever an attempt is made to determine the amount of "bandwidth" (speed of transmission) required by an application. The results of this survey indicate that there is no clear average either in the amount of emissions data transferred between organizations, or the frequency at which it is

transferred. In fact, both volume and frequency varied widely. However, several important patterns can be extracted from the data received. Most emissions inventory participants are transferring data files that are less than 10 megabytes in size. On the other hand, the range of file sizes that appeared in the survey results varied quite significantly - from 500 bytes to “over a gigabyte.” (Almost all air quality modelers will eventually work with large data files, many of which are well over 100 megabytes in size.) The following table summarizes the results of question 9, and indicates the number of responses showing an average volume in the following ranges:

<u>Average volume range</u>	<u>Number of responses</u>
< 500K	4
500K - 5Mb	17
5Mb - 10Mb	3
10Mb - 100Mb	4
> 100Mb	1

The *mean* of the volumes being currently transmitted is clearly in the half megabyte to 5 megabyte range. One response indicated the importance of the potential savings in transmission time and capacity that can be achieved from compressing files. This response indicated that an average volume of 6 megabytes could be compressed to 100 Kbytes for transmission.

The frequency of emissions data transmission also varies widely, ranging from 1 time per year to 10 times per week. Note that even the most frequent instance of emissions data transmission (twice a day) could be considered “infrequent” by today’s networking standards, where it is not unusual for some World Wide Web servers to experience 20,000 data accesses per day. The following table summarizes the responses to question 11.

<u>Average frequency range</u>	<u>Number of responses</u>
less than 1/month	7
1/month to 1/week	8
1 to 2 times per week	8
3 to 10 times per week	3
more than 10 times per week	4

The frequency varies widely, although the majority of transmissions are made in the once per month to twice per week frequency.

AIRS plays a significant role in emissions data transfers. Many of the infrequent data transfers are batch uploads to AIRS, while many of the frequent transfers are a result of *ad hoc* requests from a wide variety of sources and the use of AIRS.

## Data Sensitivity

A majority of the survey respondents (62%) indicated that they did not electronically transfer sensitive data to other emissions inventory participants. However, a significant number of individuals (14) indicated that they did transfer sensitive data, so some form of data security should play a role in any recommendation regarding the electronic transfer of emissions inventory files.

## Hardware

The most prevalent type of hardware in use by an overwhelming margin was the IBM PC-compatible personal computer. 82% of respondents indicated that this was their only or primary source of processing power. 15% of respondents indicated a connection to an IBM mainframe, although only two respondents seemed to consider this their primary source of processing power. Other responses mentioned UNIX workstations, UNIX servers, Unisys A16, Mac Quadra, Sun SPARC 20, Novell LAN, Hewlett-Packard model 9000 series 735/125 and a HP755 workstation.

The Intel 486-series of processor was, by a wide margin, the most prevalent type of PC workstation in use by the survey respondents. Only three indicated that they were still using PCs in the 386 range, while most have 486/66s available with some indicating that they were either using, or intending to purchase, the new Pentium series of processors. It is clear that, regardless of the chosen base line data transfer medium, emission inventory PC workstations will not be short of processing power.

## Future Resources

Survey respondents displayed an undeniable willingness to upgrade their data transfer capabilities *as a direct result of the EIIP*. The majority declared that they would purchase, in order of popularity:

- faster PCs or workstations (79%)
- high speed modems (71%)
- direct access to the Internet (69%)
- new software for data transfer (64%) or,
- dedicated phone lines (56%)

A barely significant number (35%) said that they would be willing to purchase new software for data formatting (e.g., X12 translators). Only 26% of respondents indicated a willingness to purchase access to a Value-Added network such as Prodigy or CompuServe.

## **Preferences**

Although a majority of the respondents replied that the Internet was the preferred method of emissions inventory data transfer, the vote was not unanimous nor without a significant range of responses. One of the salient points raised in the comments was that very large files (over 100M) may be transmitted more efficiently using tapes (in overnight mail, if time is of the essence.) Transfers using diskettes ("sneaker net") are still popular. Data transfer via modem also received mention.

The ASCII file type is clearly the most popular, with other file types not receiving enough responses to be considered "significant." Spreadsheets and word processors will obviously continue to be used for storing or formatting emissions data, as will generic text files. When comparing "CURRENT" to "PLANNED" use, the number of write-in comments regarding database files increased by one mark (from eight to nine). Popular databases (Paradox, dBase, Foxpro, etc.) were also well represented in the preferred record format results of question 17. In fact, the number of write-ins for "Other" in this section increased by 3 (from 12 to 15) over the same results from the current record format responses to question 8. The number of individuals who replied that AIRS was the preferred record format decreased by 6 (from 25 to 19) when compared with the responses to question 8, of those who currently use AIRS format.

## SUMMARY OF CONCLUSIONS: EIIP/DMC DATA TRANSFER SURVEY

The items below summarize point by point the significant conclusions from the analysis above. These conclusions will then be used to develop tentative recommendations for the next section.

### I. Media

- A. 87% of respondents currently use paper for transfer of emissions inventory data.
- B. Although about 30% of respondents currently using paper plan to eliminate paper usage, paper will still remain a major medium for transferring emissions inventory data over the next two years.
- C. 77% of respondents currently use floppy disks for data transfer, and the number of users planning to use floppies continues at this level.
- D. Magnetic tape is not used widely by respondents (25%), and use is not expected to increase significantly, although usage for very large volumes of data is significant.
- E. 74% of respondents plan to have high speed modems for data transfer. The most common current speeds are 9600 and 14.4 Kbs, though the majority of respondents plan to upgrade to 28.8 Kbs.
- F. 49% of respondents currently have direct access to Internet and 74% plan to have access in the next two years.

### II. Data Formats

- A. ASCII is the predominant file-type, with no other type showing significant usage among respondents.
- B. File formats are split equally between spreadsheets and word-processors (about 54% each *current file format* and 43% each *preferred as a file format*). The predominant spreadsheet formats are Lotus, EXCEL and QuatroPro. The word processor formats of respondents are either WORD or Word Perfect.
- C. Generic text files are a preferred file format for 33% of respondents.
- D. There was a significant write-in by respondents for preferred database file formats including Foxpro, dBase, MS Access and Paradox.
- E. 64% of respondents currently use AIRS record format, and 49% indicated AIRS as a preferred record format.
- F. 26% of respondents currently use EPS-Input and 21% indicated it as a preferred record format.



- G. 15% of respondents use EPS-Output and 15% indicated it as a preferred record format.
- H. There was very little recognition among respondents of the X.12 ANSI EDI standard as a current or preferred record format.
- I. Other application-specific record formats were strongly represented with about 31% current use and 38% preferred format responses.

### **III. Volumes and Frequency of Data Transmission**

- A. Volumes and frequency of data transfer varied widely among respondents.
- B. The *mean* of “volumes being transmitted” is in the half megabyte to 10 megabyte range. Only 13% of respondents indicated an average transmission range that exceeded 10 Mbytes. Only 10% of respondents indicated an average transmission range that was less than 500K.
- C. The frequency of transmissions for the majority of respondents is in the once per month to twice per week range. Only 18% of respondents transferred data less than once per month. Only 18% of respondents transferred data more frequently than twice per month.

### **IV. Transmission and Communications**

- A. Internet access is available to or planned by almost all respondents in the survey (74%)
- B. Respondents in the survey generally do not have access to, do not plan to have access to, and did not indicate a willingness to expend additional resources for access to Value-Added-Networks such as CompuServe, Prodigy, America Online, etc.
- C. Respondents in the survey using EPA NDPD SNA connections expect a decline in planned usage over the next two years from 36% to 15% of respondents.

### **V. Hardware**

- A. Almost all respondents use powerful Windows-capable PC-compatible workstations, and are willing to expend resources on more powerful PC processors.
- B. Other types of processors (Apple Mac, Mainframe, UNIX workstations, etc.) are not widely represented in the survey.

### **VI. Application Characteristics**

- A. Almost all respondents having high speed modems have access to data transfer

- software (primarily ProComm and Crosstalk) that includes *de facto* file transfer standard techniques such as Xmodem, Ymodem, Zmodem, and Kermit.
- B. A significant number of respondents (33%) transfer sensitive data, so security of the proposed solution is important.
  - C. One respondent indicated that compression of emissions inventory data in files by software for transmission can be significant in reducing data transfer volumes and hence data transmission capacity and time requirements.
  - D. Only 35% of respondents indicated that they would be willing to expend resources on new software for data formatting (e.g. X.12 translators), significantly less than those spending on other technologies, based upon EIIP recommendations.
  - E. Current investments in information systems to support AIRS, EPS, and other application-specific formats are significant and should not be abandoned hastily.